

Submission in Response to NSF CI 2030 Request for Information

DATE AND TIME: 2017-04-05 16:57:31

PAGE 1

REFERENCE NO: 300

This contribution was submitted to the National Science Foundation as part of the NSF CI 2030 planning activity through an NSF Request for Information, https://www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf17031. Consideration of this contribution in NSF's planning process and any NSF-provided public accessibility of this document does not constitute approval of the content by NSF or the US Government. The opinions and views expressed herein are those of the author(s) and do not necessarily reflect those of the NSF or the US Government. The content of this submission is protected by the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (<https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>).

Author Names & Affiliations

- Michael Ferris - University of Wisconsin

Contact Email Address (for NSF use only)

(Hidden)

Research Domain, discipline, and sub-discipline

Optimization and Data Science, Computer Sciences, Mathematics, Industrial and Systems Engineering

Title of Submission

From Data to Decisions

Abstract (maximum ~200 words).

We need to develop end-to-end strategies for data collection, analysis, management, privacy, security, visualization, and decision-making. A nearly universal challenge of the 21st century is effectively using the massive amounts of data that can be collected or are already available. Advances in science and technology in this century are likely to be driven by the ability to use data meaningfully, which will require new methods to sort, analyze, and model it to detect patterns and mechanisms in the world around us. These processes define the broad field of data science, which will include an interplay of machines and humans—using the best of machine learning and human expertise to build models to address challenges such as managing energy, agricultural production, and the environment responsibly. This is an essential thrust to enable wise utilization of the explosion of diverse data available to agricultural managers, energy policy-makers, and healthcare workers, patients, and consumers in the United States and beyond.

Question 1 Research Challenge(s) (maximum ~1200 words): Describe current or emerging science or engineering research challenge(s), providing context in terms of recent research activities and standing questions in the field.

A fundamental challenge is to draw together the data science community, and extensive scientific collaborations between scientists and other communities, and enable the application of fundamental mathematical research on optimization, machine learning, and data science to practical problems. NSF has provided and leveraged large funding in data science and is also leading calls for proposals for centers of excellence in machine learning and data science, which could provide the impetus to create a University based data science centers.

The Challenge will require focusing efforts in four complementary areas that will enable Universities to become a hub for data science training and application:

- mathematical foundations, modeling, algorithms, optimization, and machine learning

Submission in Response to NSF CI 2030 Request for Information

DATE AND TIME: 2017-04-05 16:57:31

REFERENCE NO: 300

PAGE 2

- database management and high throughput processing
- privacy and security
- applications in science and technology, energy, the environment, agriculture, biomedicine, and healthcare.

Question 2 Cyberinfrastructure Needed to Address the Research Challenge(s) (maximum ~1200 words): Describe any limitations or absence of existing cyberinfrastructure, and/or specific technical advancements in cyberinfrastructure (e.g. advanced computing, data infrastructure, software infrastructure, applications, networking, cybersecurity), that must be addressed to accomplish the identified research challenge(s).

We believe it would be a mistake to apply a piecemeal mechanism, where a large number of researchers build various utilities and generate expertise for specific domains without using sophisticated analysis tools, or harness the power of statistics, mathematics and computer science to build, maintain and enhance our capabilities. Some Universities are making concerted efforts to build capabilities in this area. By coordinating our Data Analytics efforts, we will better leverage one another's contributions, help attract better faculty candidates, increase our visibility and impact on among potential industrial sponsors and affiliates, putting us in a great position to establish a new knowledge and industrial strengths in these area.

Question 3 Other considerations (maximum ~1200 words, optional): Any other relevant aspects, such as organization, process, learning and workforce development, access, and sustainability, that need to be addressed; or any other issues that NSF should consider.

Every University needs a capability to collect large amounts of data, from existing data sources, and from mining new sources. Additionally, we need a suite of tools that help pose appropriate questions of that data, and augment this by tools that do inference, learning and other data knowledge tasks. We need to develop in-house capability to turn these tools into applications, and in-house capability to visualize and summarize the things that we learn and extract.

This will need leadership and workforce development. We need to augment faculty with support groups of staff members and train the next generation of interdisciplinary data scientist practitioners. Much of the required work needs expertise that is not in the skillset (or time constraints) of a faculty members or teachers but needs a level of programming skill and familiarity with a suite of computational tools that must be fostered in (permanent) skilled support staff. This is a critical need and needs immediate investment.

Consent Statement

- "I hereby agree to give the National Science Foundation (NSF) the right to use this information for the purposes stated above and to display it on a publically available website, consistent with the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (<https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>)."